Robustness Enhancement for Noncentric Quantization-Based Image Watermarking

Soo-Chang Pei, *Fellow, IEEE*, and Jun-Horng Chen, *Member, IEEE*

***Abstract*—This paper presents a novel approach for improving the robustness of the quantization-based watermarking scheme. Unlike the conventional approach, which quantizes the host value to the center of the objective interval, the proposed scheme offers the flexibility of determining the quantized value. Given tolerable embedding-induced distortion and the host value, the quantized value can be analytically determined in a closed-form solution, instead of being fixed at the center of the interval. To objectively compare the robustness, different watermarking schemes are implemented in the same distortion-to-noise ratio (DNR) scenario and then the probabilities of error detection are measured. Simulation results indicate that the proposed scheme, compared to the conventional quantization watermarking, could reduce the error probability by 0.113 when DNR is set to 0 dB, whereas the reduction is only 0.037 for the recent Wu's approach. Furthermore, without the finite possible values of the watermarked data, the proposed scheme provides sufficient nondisclosure for the quantization step, whereas in the conventional quantization-based watermarking, the quantization step can be easily inferred from the watermarked data.**

***Index Terms*—Blind detection, DNR, informed detection, quantization-based watermarking, robustness.**

I. Introduction

FOR various applications [1], the image watermarking system is developed to embed additional information into the host image. The embedded message may carry the ownership or secret information. Conventional cryptographic systems only allow holder of a valid key to access the encrypted data. However, the decrypted data became untrackable, and the copyright protection thus becomes invalid. A watermarking system can be considered a complement of the cryptographic system. Especially, the watermarked image is visually identical to the host image.

Fig. 1 illustrates a generic diagram of a watermarking system. The host signal $\mathbf{x}$ is modified to become the watermarked signal $\mathbf{s}$ to convey the watermark message $\mathbf{m}$. Generally, the watermarked signal inevitably suffers the channel effect, which may result from malicious attacks or unintentional processes. However, the extracted watermark $\hat{\mathbf{m}}$ is expected to be as close to $\mathbf{m}$ as possible. The scrambling and descrambling processes can provide three functions of the watermarking system: *security*, *cropping-resisting* [2], and *message shaping* [3].

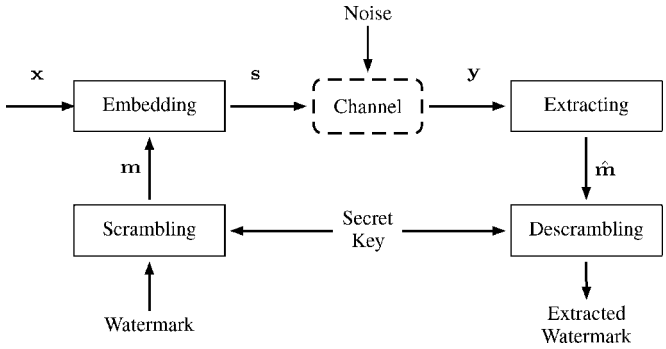

Fig. 1. Schematic diagram of a watermarking system.

Security: The scrambling and descrambling processes are generally controlled by a random seed which can be regarded as a secret key. Only those who hold the key can view the watermark message.

Cropping-resisting: If the watermark is a bilevel and visually recognizable image (like a seal, emblem, etc), the two-dimensional random permutation of the watermark image will disturb its spatial relation, preventing pirates from easily removing the watermark message by cropping.

Message shaping: The distribution of the binary watermark message is generally assumed to be equally probable, $P_r\{\mathbf{m}=0\}=P_r\{\mathbf{m}=1\}=0.5$. This is intended to remove the influence of message distribution. Therefore, via an exclusive OR process with a random message $\mathbf{z}$, where $P_r\{\mathbf{z}=0\}=P_r\{\mathbf{z}=1\}=0.5$, the distribution of the watermark message is guaranteed to be equally probable.

Actually, in the above functions, the receiver only requires the random seed, which is usually not too long, to use as the secret key to reveal the watermark message. In [4], the security requirement is achieved by sending a secret key with the same length as the watermark message. However, the watermarking scheme becomes impractical when too much overhead information is involved.

For *invisible watermarking*, the embedding should not be noticeable when constricting the extent of the superimposed distortion. Meanwhile, the watermarked image is subject to the channel effect, which may cause the wiping off of the embedded message. Accordingly, the embedded message should be sufficiently robust to survive reasonable[1] attacks on the

Manuscript received Ocotober 17, 2004; revised May 9, 2006. This work was supported in part by the National Science Council, Taiwan, R.O.C., under Contracts NSC 94-2213-E-002-072 and NSC 93-2752-E-002-006-PAE. This paper was recommended by Associate Editor E. Izquierdo.

S.-C. Pei is with the Department of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C. (e-mail: pei@cc.ee.ntu.edu.tw).

J.-H. Chen is with the Department of Communication Engineering, Oriental Institute of Technology, Taipei 220, Taiwan, R.O.C. (e-mail: jhchen@mail.oit.edu.tw).

Digital Object Identifier 10.1109/TCSVT.2006.885174

[1]Generally, an applicable watermarking scheme ensures that the severe degradation of fidelity of the watermarked image follows the attack which makes the embedded message be irretrievable. Namely, the attacks, which seriously deteriorate the visual quality of the watermarked image, are excluded from the discussion of the watermarking scheme.

watermarked image. Generally, the embedded signal with a higher energy strength makes the watermarked image more capable of withstanding attack, yet induces more perceptual degradation. Therefore, the designers of watermarking system usually face the dilemma of robustness and transparency.

Additionally, the quantity of embedded information (namely the capacity), which determines the ability to distinguish different messages (or different proprietors), is also a fundamental problem of watermarking schemes. Usually, capacity is measured based on the number of bits which can be embedded in a host sample (*bits per sample*). How many bits of the embedded message are satisfactory depends on the application the watermarking scheme is applied to. In the application of access control, capacity demand is not so much [5]. However, the applications of fingerprinting and information-hiding may demand high payloads of the embedded information. Recently, many researches of the digital rights management (DRM) addressed the issue of benchmark of the watermarking system in multidimensional criteria evaluation. The optimization of the watermark parameters (robustness, transparency, and capacity) is mutually competitive and cannot be clearly done at the same time [6]. A reasonable method of comparing different watermark schemes involves measuring the error probabilities of the messages extracted from the watermarked images with the same DNR,[2] where the watermarked images contain equal quantity of information. Naturally, unlimited embedding-induced distortion can not be allowed in practical applications. The measurement and comparison in this work consider the watermarked image should be indistinguishable from the host image.

Cox *et al.* [7] proposed a simple and extensively used scheme, *spread spectrum* (SS), which purely adds a watermark signal $\mathbf{w}(m)$ to the host signal $\mathbf{x}$

$$\mathbf{s} = \mathbf{x} + \mathbf{w}(m) = \mathbf{x} + \delta(m)\mathbf{v} \tag{1}$$

where $\delta(m)$ denotes a scalar function of embedded message m, and $\mathbf{v}$ represents a pseudo-random spreading vector. In [7], the host signal $\mathbf{x}$ is a vector in the frequency domain of the host image. If the host signal $\mathbf{x}$ is known to the receiver (i.e., informed detection [8]), the embedded message can be easily and accurately estimated from the attacked watermarked signal $\mathbf{y} = \mathbf{s} + \mathbf{n}$, because the host-interference is removed and the watermark signal $\mathbf{w}(m)$ is uncorrelated with the additive noise $\mathbf{n}$. However, in some applications, the host signal is unlikely to be available at the receiver (blind detection). Malvar *et al.* [9] thus proposed the *improved spread spectrum* (ISS), which modulates the energy of the inserted watermark to compensate for the host-interference. This scheme modifies (1) as

$$\mathbf{s} = \mathbf{x} + \delta(\mathbf{x}, m)\mathbf{v}. \tag{2}$$

A simpler version of $\delta(\mathbf{x}, m)$ is a linear function, that is, $\delta(\mathbf{x}, m) = \eta m - \lambda \mathbf{x}^t \mathbf{v}$. However, the later section demonstrates that the ISS is beneficial only for high embedding-induced distortion or low capacity.

Chen and Wornell [10] proposed *quantization index modulation* (QIM) which has been demonstrated to be an effective method for digital watermarking because it could blindly estimate the embedded message with almost the same accuracy as the accuracy of the host image being available at the receiver. Especially, *spread-transform dither modulation* (STDM) is a low-complexity realization, in which the projection of the host signal along a pseudo-random vector is quantized to the center of the nearest interval referring to the embedded message. Generally, the quantization intervals are nonoverlapping and all have the same size. Thus, all of the quantized values are equally separated in the range of the possible values and are alternately assigned to messages "1" and "0." To enhance the security protection, Wu [11] recently proposed that the *look-up table* (LUT) should be built on top of a general quantization-based watermarking, in which the quantization intervals are randomly assigned to "1" or "0," with the constraint that the runs of "1" and "0" have limited length. However, the approach developed by Wu can only improve the robustness of QIM in the low DNR case, where the probability of error detection is usually high.

Many adaptive watermarking methods were proposed for optimizing the performances of robustness and transparency simultaneously. Barni *et al.* [12] relied on some of the ideas exposed in [7] and exploited the spatial masking characteristics of the HVS to enhance the transparency property. The strength of the watermark embedding is modulated according to the local spatial statistics. For the highly textured regions, which are characterized by low-noise sensitivity, the watermark can be embedded to a mighty extent, whereas in the uniform regions, which are more sensitive to change, the watermark is embedded only to a minor extent. Different from the general wavelet-based watermarking, Bao and Ma [13] applied the QIM [10] approach on singular values of the wavelet domain for image authentication. On the basis of the idea that slight variations of singular values do not affect the visual perception of the image, the watermark bits are embedded on the singular values for each of the wavelet domain blocks, with the quantization steps adaptive to the statistical model of the blocks. Chang *et al.* [14] adopted the fuzzy adaptive resonance theory (Fuzzy-ART) classification to identify appropriate DCT blocks for watermarking. The embedding strength is determined by the intensities of the high and low frequency components in the selected blocks. Stronger embedding strength is used in those blocks with fewer large magnitude DCT coefficients and vice versa.

Following the conventional quantization technique, which minimizes the quantization error to satisfy the needs of coding or compression, most of the quantization-based watermarking schemes quantize the host value to the center of the objective interval. This work proposed that the quantized value is no longer located at the center of the quantization interval, but rather is determined by the host value and the embedded message. This flexibility of the quantized value diminishes the embedding-induced distortion, and can coordinate the transparency and robustness requirements of a watermarking

[2]the ratio of the embedding-induced distortion and the power of the attacking induced noise.

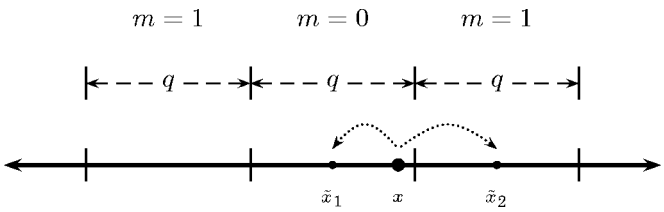

Fig. 2. Conventional quantization-based watermarking.

scheme. Experimental results demonstrates that the proposed scheme effectively improves the robustness performance of other quantization-based watermarking schemes.

Once the robustness is improved, the watermarking approach can be more extensively applied to various scenarios. Especially in the applications of data hiding, lower error rate implies wider transmission bandwidth and more amount of conveyable information. For example in video watermarking, watermark annotation makes the host data convey side information for achieving additional functionalities, automatic monitors of the broadcast data, and device control, etc.

II. Quantization-Based Watermarking

The quantization-based watermarking scheme is discussed here to demonstrate the performance improvement of the proposed scheme. The quantization-based watermarking schemes are shown to be able to extract the embedded message without requiring the host signal. Additionally, although the spread spectrum (SS) [7] based watermarking is proposed to be suitable for informed detection, Malvar *et al.* [9] declared that blind detection can also be applicable to the ISS watermarking scheme. Therefore, this section also presents the performance analyses of SS and ISS in order to compare them with the proposed scheme in the following sections.

A. Conventional Quantization-Based Watermarking

Fig. 2 illustrates the main concept of the quantization-based watermarking. The embedding domain is divided into subsets with quantization step q, and the subsets are alternately designated as "interval-0". Let x denote the host value and $m = \{0, 1\}$ represent the one-bit message to be embedded, then the embedding is denoted as

$$s = Q\left[x + \delta(m)\frac{q}{2}\right] - \delta(m)\frac{q}{2} \tag{3}$$

where $Q[\cdot]$ is a uniform quantizer with step size $2q$, and $\delta(m)$ denotes a message modulating function mapping the embedded message to $\{1, -1\}$.

Inheriting from the quantization techniques applied to compressing or coding, which attempt to minimize the quantization error, the quantization in (3) chooses the center of the interval as the quantized value. However, the quantization-based watermarking should additionally consider the message to be embedded. As shown in Fig. 2, both $\tilde{x}_1$ and $\tilde{x}_2$ are the centers of the corresponding intervals. In the embedding process, the host value x is replaced by $\tilde{x}_1$ when x is located in the interval which represents the same message as that to be embedded; and $\tilde{x}_2$ is chosen as the quantized value in the opposite case. If the above two cases occur with equal probability, the embedding-induced distortion, measured by *mean squared error* (MSE), is

$$D = E\left[||s - x||^2\right] = \frac{1}{2}\int_0^{\frac{q}{2}} \left[x^2 + (q - x)^2\right] f_{\mathbf{x}}(x)dx \tag{4}$$

where $f_{\mathbf{x}}(x)$ denotes the density function of x in the region of 0 to $q/2$. Assuming x is uniformly distributed over the region,[3] the distortion D is expressed by

$$D = \frac{q^2}{3}. \tag{5}$$

Regarding the extracting process, the receiver simply determines the embedded message according to the interval in which the received value is located. Notably, the receiver does not require the host signal. However, the channel noise n, which may result from malicious attacks or unintentional signal processes, can contaminate the watermarked signal s such that the received signal is $y = s + n$. The extracted message bit $\hat{m}$ is correct provided y and s are located in the intervals which represent the same message. Namely, the detection error will occur when the strength of the attack noise n is in the following range:

$$\left\{\cdots \bigcup \left(-\frac{7q}{2}, -\frac{5q}{2}\right) \bigcup \left(-\frac{3q}{2}, -\frac{q}{2}\right) \bigcup \left(\frac{q}{2}, \frac{3q}{2}\right) \bigcup \left(\frac{5q}{2}, \frac{7q}{2}\right) \bigcup \cdots\right\}$$

Therefore, the error probability of detection is

$$\begin{aligned} P_e &= \int_{\frac{q}{2}}^{\frac{3q}{2}} f_{\mathbf{n}}(n)dn + \int_{-\frac{3q}{2}}^{-\frac{q}{2}} f_{\mathbf{n}}(n)dn \\ &\quad + \int_{\frac{5q}{2}}^{\frac{7q}{2}} f_{\mathbf{n}}(n)dn + \int_{-\frac{7q}{2}}^{-\frac{5q}{2}} f_{\mathbf{n}}(n)dn + \cdots \\ &\approx 2\int_{\frac{q}{2}}^{\frac{3q}{2}} f_{\mathbf{n}}(n)dn \end{aligned} \tag{6}$$

where $f_{\mathbf{n}}(n)$ denotes the density function of the noise n. The approximation of (6) is only valid when $f_{\mathbf{n}}(n)$ is symmetrical and unimodal.[4] If the noise n is the Gaussian distribution with zero mean and variance σ_n^2, (6) becomes

$$\begin{aligned} P_e &= \operatorname{erfc}\left(\frac{q}{\sqrt{8}\sigma_n}\right) - \operatorname{erfc}\left(\frac{3q}{\sqrt{8}\sigma_n}\right) + \operatorname{erfc}\left(\frac{5q}{\sqrt{8}\sigma_n}\right) \\ &\quad - \operatorname{erfc}\left(\frac{7q}{\sqrt{8}\sigma_n}\right) + - \cdots \\ &\approx \operatorname{erfc}\left(\frac{q}{\sqrt{8}\sigma_n}\right) \end{aligned} \tag{7}$$

[3]The host value x is not always uniformly distributed. However, if the quantization step q is sufficiently small, the distribution of x can be approximately considered uniform.

[4]Certainly, the mode of the density function will be zero. Besides, the rate of increase below the mode and the rate of decrease above the mode should be as fast as possible, otherwise the approximation will yield considerable inaccuracy.

where the complementary error function $\mathrm{erfc}(\cdot)$ is defined as

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt.$$

Clearly, (6) and (7) are influenced by quantization step q. Especially in (7), larger value of q reduces the error probability, yet increases the embedding-induced distortion. Therefore, the DNR γ is generally defined as

$$\gamma \triangleq \frac{D}{\sigma_n^2}. \tag{8}$$

Thus, the probability of error detection can be represented in terms of γ

$$\begin{aligned} P_e &= \mathrm{erfc}\left(\sqrt{\frac{3\gamma}{8}}\right) - \mathrm{erfc}\left(\sqrt{\frac{27\gamma}{8}}\right) + \mathrm{erfc}\left(\sqrt{\frac{75\gamma}{8}}\right) - + \cdots \\ &\approx \mathrm{erfc}\left(\sqrt{\frac{3\gamma}{8}}\right). \end{aligned} \tag{9}$$

Though (9) fairly describes the robustness performance, the embedding-induced distortion is not unlimited. After all, the watermarking is in vain if the transparency requirement can not be achieved.

Chen and Wornell [10] named the embedding technique presented in (3) *dither modulation* (DM), which is derived from *dither quantization* [15]. The dither quantization attempts to obtain a higher reproduction quality from the quantized signal than the general quantization. However, the dither signal used in dither quantization is a random signal, yet is a deterministic signal determined by the embedded message in watermarking. The DM approach also appears to resemble the so-called least-significant-bit(s) (LSB) [16] coding or low-bit(s) modulation (LBM), but in reality does not. The LSB coding or LBM begins by setting the least-significant-bit(s) of the host value x to zero. One of two values, represented as "1" and "0," is added to x according to the value of the embedded message. These two representative values are usually the minimum and maximum of the range of the least-significant-bit(s). Fig. 3 shows the input-output (x versus s) characteristic functions of DM and LBM coding, respectively. Since the quantized values for $m = 1$ and $m = 0$ in Fig. 3(a) and (b) are all separated by a distance of q, they have the same abilities to withstand the noise attack. Nevertheless, the DM and LBM coding approaches pay different costs of embedding-induced distortion. In Fig. 3, the area enclosed between the characteristic function and the 45° line indicates the *absolute distortion*, $|s - x|$, resulting from the embedding process. Observing the range for $x = -q$ to q, the average area for embedding message $m = 1$ and $m = 0$ is q^2 by DM, and $1.25q^2$ by LBM coding. That is, to achieve the same robustness performance, the DM approach pays less cost in terms of embedding-induced distortion than LBM coding. The MSE of the LBM coding can also be calculated by

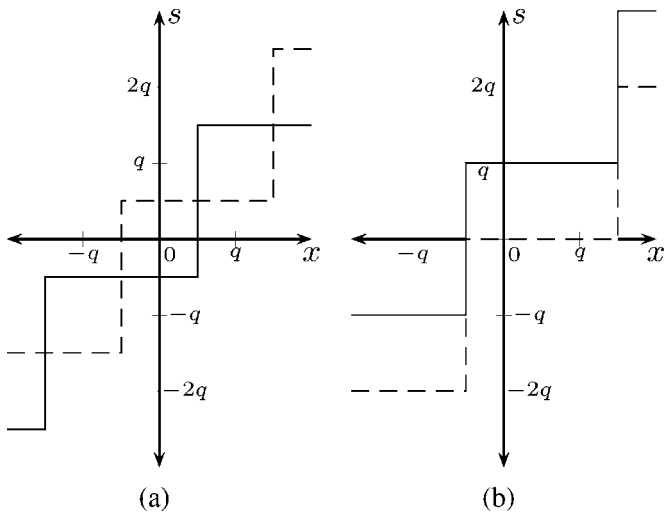

Fig. 3. Input-output characteristic functions of (a) dither modulation and (b) LBM coding. The solid line is for $m = 1$ and the dashed line is for $m = 0$.

$$D = \frac{1}{2q} \int_{-\frac{q}{2}}^{\frac{3q}{2}} \frac{[(q-x)^2 + x^2]}{2} dx = \frac{7}{12} q^2. \tag{10}$$

Comparing (5) and (10), the LBM coding induces 1.75 times more MSE than the DM approach.

To alleviate the embedding-induced distortion superimposed on one sample of the host signal, the host signal can be partitioned as the L-dimensional vectors $\mathbf{x}$, each vector is then transformed to a scalar x which is the embedding object. The embedding domain comprises all possible values of x. Consequently, the capacity of the watermarked signal is $1/L$ bits per sample. By embedding the message into transformed value x, the distortion is spread over each dimension of $\mathbf{x}$. For example, in [10], the authors projected the host vectors $\mathbf{x}$ along randomly determined directions and quantized the projection values, and named this method *spread transform dither modulation* (STDM). Besides, the randomness of the directions to be projected on can provide a secret communication between the embedder and the extractor. In STDM, the embedding-induced distortion is $1/L$ of (5)

$$D = \frac{q^2}{3L}. \tag{11}$$

The host signal is not always partitioned into vectors to reduce the distortion on one sample. In [11], the author traded robustness for the visual quality of the watermarked image. Regarding the increase in error probability, each bit of the watermark message was embedded repetitively. Using the repetition coding method can reduce the overall probability of detection error. STDM and repetition coding are compared in Section II-B.

In the conventional quantization-based watermarking, the partitioned intervals are alternately designated "interval-1" and "interval-0." In [11], the author recently introduced a LUT approach instead of following this convention. A predefined table, which randomly maps the quantization intervals to message

bits "1" or "0," enhances security of the watermarking system. Compared to the conventional quantization-based watermarking for the same quantization step size, the LUT approach results in more embedding induced distortion and less error probability. However, the author of [11] demonstrated that the LUT approach outperforms conventional quantization-based watermarking in robustness in the same DNR scenario. If runs of "1" and "0" are limited to a length of two, then the overall embedding-induced distortion is shown to be [11]

$$D = \frac{q^2}{2} \tag{12}$$

and the probability of error detection under the Gaussian noise then can be approximated by

$$P_e = \frac{2}{3}\mathrm{erfc}\left(\frac{q}{\sqrt{8}\sigma_n}\right) = \frac{2}{3}\mathrm{erfc}\left(\frac{\sqrt{\gamma}}{2}\right). \tag{13}$$

Actually, the LUT approach only achieves an improvement in the case of low DNR. Undoubtedly, the error probability increases as the DNR decreases. That is, under stringent detection accuracy requirements, the application of the LUT approach is confined to the cases of low capacity. This study shows that the proposed scheme outperforms the other quantization-based watermarking schemes in most of the applicable scenarios, and hence raises the feasibility of embedding the information with high capacity.

B. Spread Transform Versus Repetition Coding

As discussed in Section II-A, either spread transform or repetition coding can be used to trade capacity for reducing embedding-induced distortion. This investigation compares the contributions of these two approaches to the quantization-based watermarking schemes. The comparison is based on the premise that two approaches are applied to embed the same amount of information, and the probability of error detection for the case of the same DNR is then presented, where the attack noise is Gaussian. This investigation assumes that, in the spread transform approach, the host signal is partitioned into L-dimensional vectors, each vector conveys one-bit of the message, and the same message bit is repeatedly embedded into L samples of the host signal in the repetition coding.

From (7), (8), and (11), the probability of error detection for the spread transform is

$$P_{e,\mathrm{ST}} = \mathrm{erfc}\left(\sqrt{\frac{3\gamma L}{8}}\right). \tag{14}$$

For repetition coding, the decoding process should incorporate the majority voting. That is, the overall probability of detection error will be

$$P_{e,\mathrm{RC}} = \sum_{k>L/2} \binom{L}{k} p^k (1-p)^{L-k} \tag{15}$$

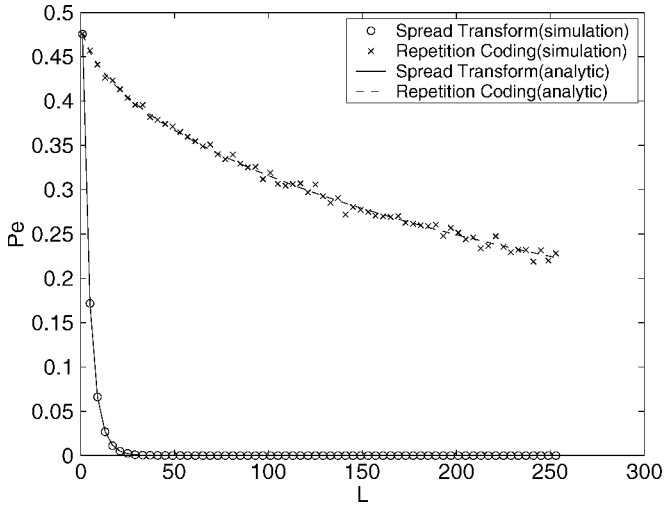

Fig. 4. Comparison of the contributions to quantization-based watermarking for spread transform and repetition coding.

where

$$p = \mathrm{erfc}\left(\sqrt{\frac{3\gamma}{8}}\right).$$

To avoid the ambiguity associated with majority voting, L is set to an odd number ranging from 1 to 255. Regarding the setting of DNR, a practical condition is considered below.

The peak signal-to-noise ratio (PSNR) is most commonly adopted to assess the image quality, and is defined as

$$\mathrm{PSNR} \triangleq 10 \cdot \log_{10}\left(\frac{255^2}{\epsilon}\right) \tag{16}$$

where ϵ denotes the MSE superimposed on one sample. Moreover, let PSNR_w and PSNR_a represent the PSNRs of the watermarked and the attacked watermarked images, respectively. The corresponding DNR can be determined as

$$\begin{aligned}\mathrm{DNR} &= 10 \cdot \log_{10}\left(\frac{\epsilon_w}{\epsilon_a - \epsilon_w}\right) \\ &= 10 \cdot \log_{10}\left(\frac{1}{10^{\left(\frac{\mathrm{PSNR}_w - \mathrm{PSNR}_a}{10}\right)} - 1}\right).\end{aligned} \tag{17}$$

For example, the DNR is set to -3.35 dB to reflect the situation where the PSNR of the watermarked and the attacked watermarked images are 40 and 35 dB, respectively.

Fig. 4 illustrates the analytical results of (14) and (15), and the experimental results for 1 000 000 randomly generated samples, with the DNR set to 3. The simulation results are clearly consistent with the analytical results. Both the spread transform and repetition coding can lower the detection error by embedding less information. Furthermore, the spread transform offers a greater contribution to the improvement of robustness than repetition coding.

C. Spread Spectrum (SS) Watermarking

The original spread spectrum watermarking [7], unlike the quantization-based watermarking, is difficult to be applied to cases involving blind detection. In the quantization-based watermarking, the detection could be guaranteed to be error-free in the absence of noise even when the host signal is unavailable, but not in the SS watermarking. The correlation-base detection, which is the most common method of extracting the message embedded by SS watermarking, has the weakness of combating the interference from the host signal.

As discussed in Section I, let the watermarked signal be denoted by $\mathbf{s}=\mathbf{x}+\delta(m)\mathbf{v}$ and the attacked watermarked signal be denoted by $\mathbf{y}=\mathbf{s}+\mathbf{n}$. If the components of the spreading vector $\mathbf{v}$ are randomly drawn from $\{\sigma_v,-\sigma_v\}$, the embedding-induced distortion is

$$D=L\sigma_v^2 \tag{18}$$

where L denotes the dimension of the spreading vector, and hence the capacity of embedding is $1/L$ bits per sample. Regarding the robustness analysis, the probability of detection error, under zero-mean Gaussian noise attack, is given by [9]

$$P_e=\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{L\sigma_v^2}{2\left(\sigma_x^2+\sigma_n^2\right)}}\right) \tag{19}$$

where σ_x^2 and σ_n^2 represent the variances of the host signal and noise, respectively. Moreover, the DNR γ is defined as in (8). Equation (19) is determined by the assumption that the host signal is Gaussian distributed with zero mean, and is unavailable at the receiver end. If the extracting involves the host signal (*informed detection*), without the host signal interference, the probability of error detection is

$$P_e=\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{L\sigma_v^2}{2\sigma_n^2}}\right). \tag{20}$$

Though the informed detection is of limited usefulness in practice, (20) can provide a goal which the watermarking schemes with blind detection pursue.

To alleviate the interference of the host signal in blind detection, Malvar *et al.* [9] proposed the ISS which modulates the energy of the inserted watermark to compensate for the host-interference. A simpler version of the ISS is

$$\mathbf{s}=\mathbf{x}+\left(\eta\delta(m)-\lambda\mathbf{x}^t\mathbf{v}\right)\mathbf{v}. \tag{21}$$

In [9], the authors proposed a method of tuning the parameters η and λ so as to cause the same degree of distortion as SS and minimize the probability of detection error. Thus, the bit-error rate is given by

$$P_e=\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{L\sigma_v^2-\lambda^2\sigma_x^2}{2\left(\sigma_n^2+(1-\lambda)^2\sigma_x^2\right)}}\right) \tag{22}$$

where

$$\lambda=\frac{1}{2}\left[\left(1+\frac{\sigma_n^2}{\sigma_x^2}+\frac{L\sigma_v^2}{\sigma_x^2}\right)-\sqrt{\left(1+\frac{\sigma_n^2}{\sigma_x^2}+\frac{L\sigma_v^2}{\sigma_x^2}\right)^2-4\frac{L\sigma_v^2}{\sigma_x^2}}\right]. \tag{23}$$

D. Robustness Improvement of the Previous Works

This investigation presents and discusses the improvements in robustness achieved by previous works. As the applicable condition adopted in Section II-B, the PSNR of the watermarked and the attacked watermarked images are set to 40 and 35 dB, respectively. Fig. 5 shows the probabilities of error detection for different ranges of L. The number of bits that can be embedded into the host image decreases with increasing L. Besides, in the SS-based watermarking, the σ_x is set to be 45.[5]

Fig. 5(a) clearly illustrates that the ISS approach indeed improves the robustness in the SS approach, especially in cases involving low capacity. However, the ISS approach can only outperform the quantization-based watermarking in cases involving very low capacity, as shown in Fig. 5(b). Notably, the 10-based logarithms of the low probabilities are presented to demonstrate their performances clearly. However, few bits can be embedded into an image in such low-capacity cases. For example, only 64 bits of a message can be embedded into an image of 256×256 for $L=1024$.

Regarding the LUT approach, Fig. 5(c) shows that it has lower error rate than the conventional quantization-based watermarking in high-capacity cases, which have very high P_e. Namely, the LUT approach is suitable for application in scenarios which do not demand high accuracy or capacity. The following section presents the proposed scheme which achieves superior robustness to the conventional quantization-based watermarking in most of the applicable scenarios.

III. Improved Quantization-Based Watermarking

For the conventional quantization-based watermarking illustrated in Fig. 2, the watermark message is embedded by modifying the host value x to the quantized value $\hat{x}_1$ or $\hat{x}_2$, which is the center of the objective interval. Choosing the center as the quantized value is inherited from the quantization approach applied to compression or coding in order to minimize the quantization error. The rationale for the proposed scheme is that the watermarking does not demand a minimal and limited number of quantized values to lower the number of bits required for

[5]From (19), σ_x influences the P_e in spread spectrum watermarking. Therefore, the value of 45 is chosen for mimicking the host image is "Lena" which is a common test image.

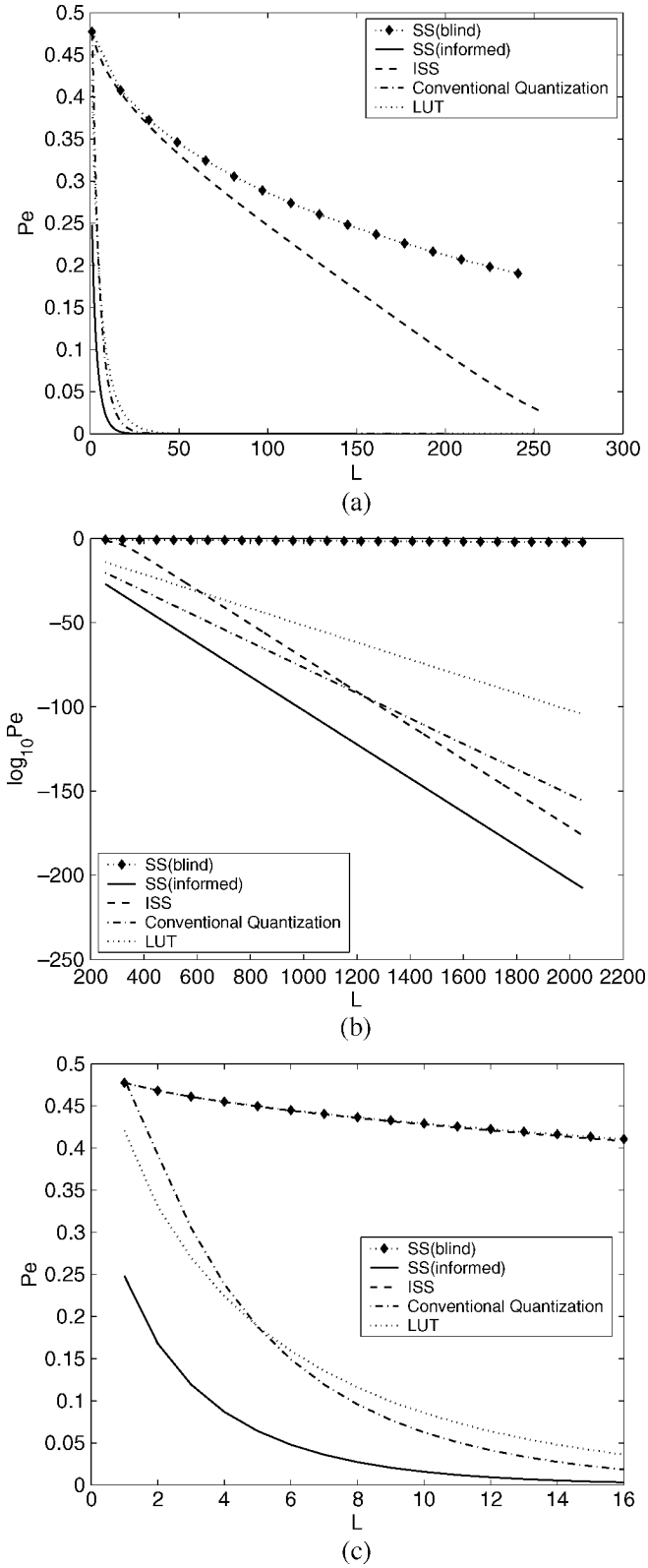

Fig. 5. Comparison of various approaches. (a) $L = 1 \sim 255$. (b) $L = 256 \sim 2048$. (c) $L = 1 \sim 16$.

storage or transmission in compression or coding. Accordingly, the quantized values should not be fixed at the centers of the intervals, but rather should be determined flexibly with reference

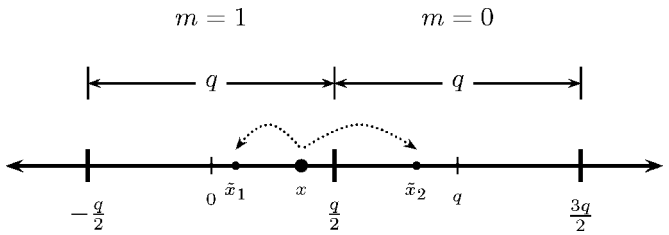

Fig. 6. Quantized values are determined according to the embedded message and the host feature x.

to the host value and considering the embedding-induced distortion and robustness.

As shown in Fig. 6, let the host feature $x = \alpha q, 0 < \alpha < 1/2$, and the quantized values $\tilde{x}_1 = \beta_1 q, 0 < \beta_1 < 1/2$, and $\tilde{x}_2 = \beta_2 q, 1/2 < \beta_2 < 1$, selected according to the messages $m = 1$ and $m = 0$, respectively. The embedding-induced distortion will be

$$D = \frac{q^2}{2}\left[(\alpha - \beta_1)^2 + (\alpha - \beta_2)^2\right] \tag{24}$$

if the distribution of the binary watermark message is equally probable.

Like (6), for the symmetrical and unimodal noise n, the probability of detection error will be approximately

$$P_e = \frac{1}{2}\left[\int_{-(\frac{1}{2}+\beta_1)q}^{-(\frac{3}{2}+\beta_1)q} f_{\mathbf{n}}(n)dn + \int_{(\frac{3}{2}-\beta_1)q}^{(\frac{1}{2}-\beta_1)q} f_{\mathbf{n}}(n)dn + \int_{-(\beta_2+\frac{1}{2})q}^{-(\beta_2-\frac{1}{2})q} f_{\mathbf{n}}(n)dn + \int_{(\frac{3}{2}-\beta_2)q}^{(\frac{5}{2}-\beta_2)q} f_{\mathbf{n}}(n)dn\right]. \tag{25}$$

Fig. 7(a) and (b) reveals that the two-dimensional contour plots of D and P_e, respectively, where the noise n is assumed to be Gaussian. In Fig. 7(a), the host value x is $0.3q$, for quantization step q of 3, while in Fig. 7(b), the error-rate is found by (25) regardless of the host value. Clearly, the distortion D and error rate P_e assume radiation shape increasing from $(\beta_1, \beta_2) = (\alpha, \alpha)$ and $(\beta_1, \beta_2) = (0, 1)$, respectively. Observing these two plots, modifying the host value to the points (β_1, β_2) in an iso-distorted line in Fig. 7(a) has different probabilities of error detection in Fig. 7(b); and modifying the host value to the points (β_1, β_2) in an iso-erroneous line in Fig. 7(b) also induces different extents of distortion in Fig. 7(a). Thus, multiple possible quantized values (β_1, β_2) exist for a tolerable distortion limit, and the value which causes the minimum error probability should be selected. Furthermore, for the fixed DNR, Fig. 8(a) and (b) shows that the best determination of (β_1, β_2) depends on the host value $x = \alpha q$, rather than being constantly fixed at (0,1).

Fig. 9 illustrates how to determine β_1 and β_2 for a given constraint on embedding-induced distortion, where Fig. 9(a) and (b) is shown for $\alpha > 0$ and $\alpha < 0$, respectively. From (24), all of the points (β_1, β_2) on the dash-circles centered at (α, α) in

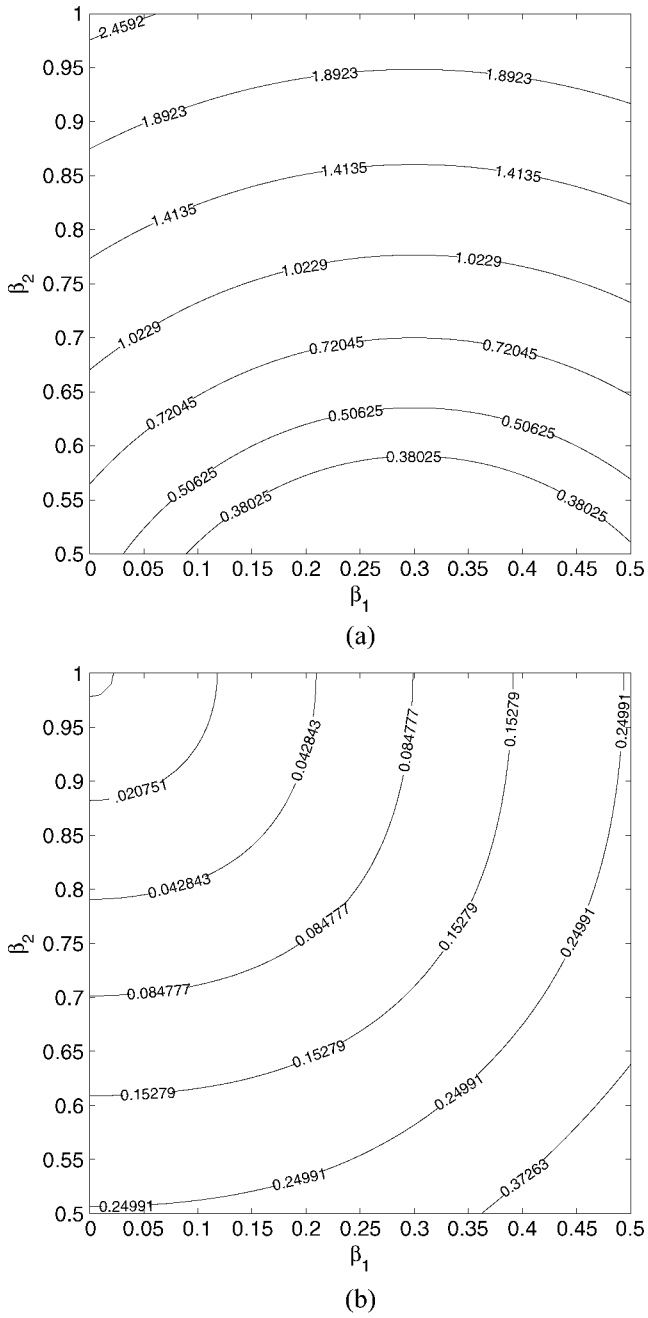

Fig. 7. Embedding-induced distortion and the error probability of detection for different β_1 and β_2. (a) D, for $q = 3$ and $\alpha = 0.3$. (b) P_e, for $q/\sigma = 5$.

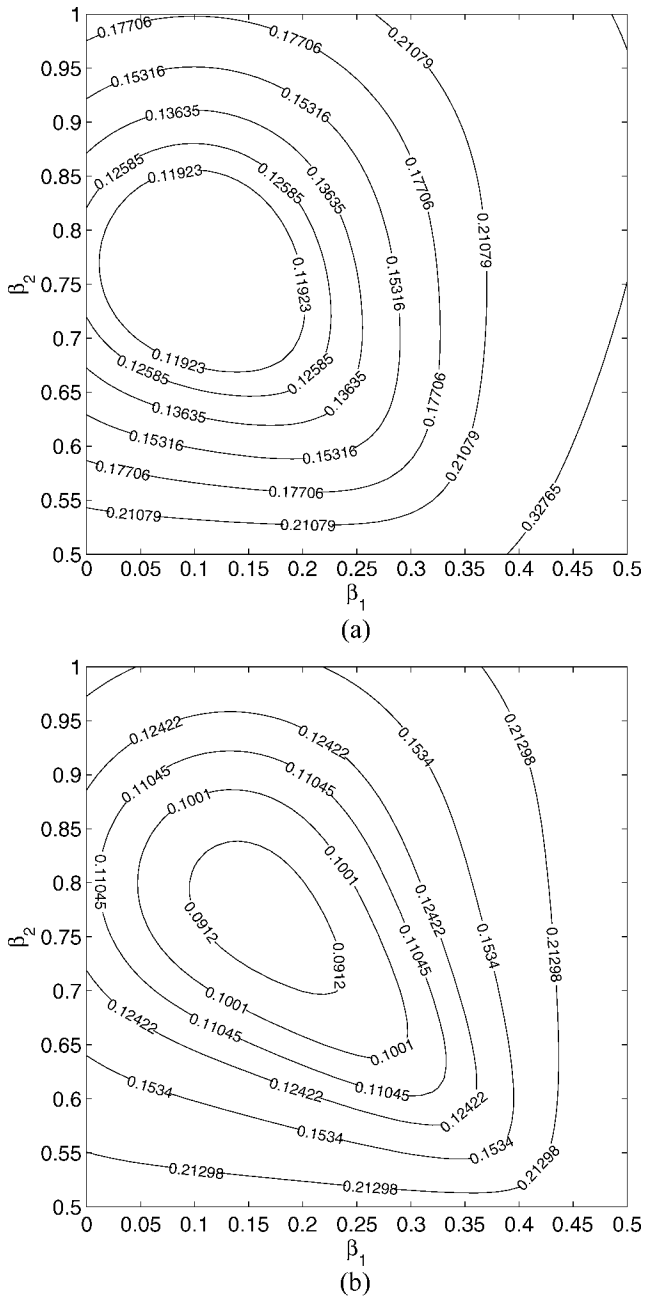

Fig. 8. Probability of error detection for different β_1 and β_2 for DNR of 3 dB. (a) $\alpha = 0.3$. (b) $\alpha = 0.4$.

Fig. 9(a) and (b) induce the same extent of distortion. Moreover, the distortion increases with increasing distance from (α, α). However, only the points in the gray zone are valid candidates for quantized values. Notably, in Fig. 7(b), the error rates radially increase from the $(\beta_1, \beta_2) = (0, 1)$ for $\alpha > 0$. Consequently, within the range of the tolerable distortion limit, the quantized values (β_1, β_2) should be as close to (0,1) as possible; and (β_1, β_2) should be as close to (0,−1) as possible when $\alpha < 0$. Based on the above observation, for $\alpha > 0$, (β_1, β_2) is proposed to be the intersection of the circle constraining the distortion and the line between (α, α) and (0,1). That is, the following equations are solved:

$$\begin{cases} (\alpha - \beta_1)^2 + (\alpha - \beta_2)^2 = R^2 \\ \beta_2 = \frac{\alpha - 1}{\alpha}\beta_1 + 1 \end{cases} \tag{26}$$

where R denotes the radius of the circle in Fig. 9(a)

$$R = \frac{\sqrt{2D_w}}{q} \tag{27}$$

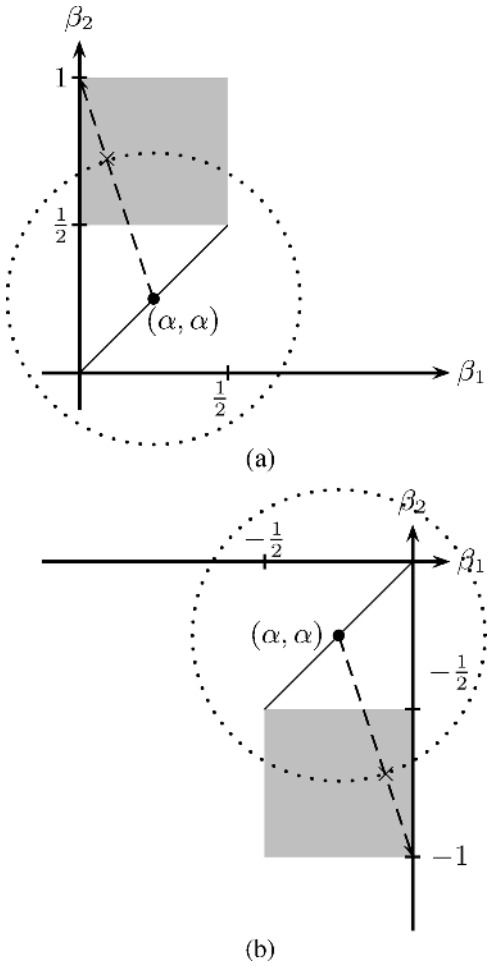

Fig. 9. Determining β_1 and β_2 for (a) $0 < \alpha < 1/2$ and (b) $-(1/2) < \alpha < 0$. (Only the points in the gray zone are valid candidates for quantized values.)

and is determined by the tolerable embedding distortion D_w and quantization step q. Since α ranges from 0 to 1/2, R should be in the range

$$\frac{1}{2} < R < \frac{1}{\sqrt{2}} \tag{28}$$

so that a valid (β_1, β_2) in the gray zone can be found. Accordingly, the quantization step q is set in the following range:

$$2\sqrt{D_w} < q < 2\sqrt{2D_w}. \tag{29}$$

The solution of (26) for $0 < \beta_1 < 1/2$ and $1/2 < \beta_2 < 1$, is given by

$$\begin{cases} \beta_1 = \alpha - \dfrac{R}{\sqrt{1+\left(\frac{\alpha-1}{\alpha}\right)^2}} \\ \beta_2 = \alpha + \dfrac{R}{\sqrt{1+\left(\frac{\alpha}{\alpha-1}\right)^2}}. \end{cases} \tag{30}$$

Similarly, for $\alpha < 0$, solving the following equations:

$$\begin{cases} (\alpha-\beta_1)^2 + (\alpha-\beta_2)^2 = R^2 \\ \beta_2 = \frac{\alpha+1}{\alpha}\beta_1 - 1 \end{cases} \tag{31}$$

results in:

$$\begin{cases} \beta_1 = \alpha + \dfrac{R}{\sqrt{1+\left(\frac{\alpha+1}{\alpha}\right)^2}} \\ \beta_2 = \alpha - \dfrac{R}{\sqrt{1+\left(\frac{\alpha}{\alpha+1}\right)^2}}. \end{cases} \tag{32}$$

Consequently, given a tolerable distortion D_w, the quantization step can be determined flexibly using (29), and the quantized values can be analytically determined in a closed-form solution according to the host value and the message to be embedded by (30) and (32), respectively.

Additionally, since the quantized values are always at the centers of the quantization intervals in the conventional quantization-based watermarking schemes, the watermarked value can be regarded as the data drawn from a finite set which consists of the points separating a multiple of quantization step from others. Thus, the quantization step is easily inferred. However, the possible values of the watermarked data of the proposed scheme are no longer finite, creating a barrier to the disclosure of the quantization step.

IV. Experimental Results

A. Gaussian Distributed Host Signal

In this subsection, the host signal comprises 1 000 000 samples drawn from the pseudo-random Gaussian distribution $\mathcal{N}(0, \sigma_x)$, and the watermark message is also randomly generated with $P_r\{m = 1\} = P_r\{m = 0\} = 0.5$, where each sample of the host signal conveys one bit of the watermark message. The watermarked signal is subjected to attacks which superimpose Gaussian noise $n \sim \mathcal{N}(0, \sigma_n^2)$ on the watermarked signal to the extent that the DNR ranges from −10 to 5 dB.

In Fig. 10, all of the quantization steps are adjusted to the extent that the embedding-induced distortion D is 6. This setting somewhat resembles the case in which the host image is an 8-bit gray-level image and the PSNR of the watermarked image is 40 dB. From [9], the performance of informed-SS could be presented as a bound for comparison. Obviously, the proposed scheme has a lower error rate than the conventional quantization watermarking scheme and the LUT approach in scenarios where the DNR ranges from −10 to 5 dB. The proposed scheme improves the robustness and approaches the bound more closely than other quantization-based watermarking schemes. If the DNR is set to 0 dB, the proposed scheme, compared to the conventional quantization watermarking, could even reduce the error probability by 0.113, whereas the reduction is only 0.037 for the LUT approach. Especially, the proposed scheme can outperform the conventional approach in cases where the LUT approach cannot.

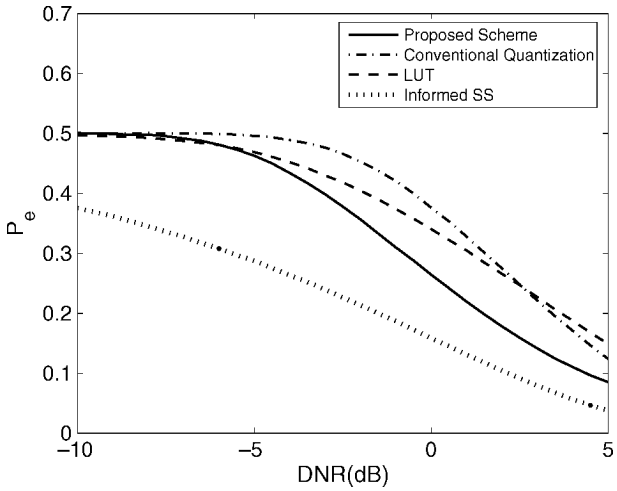

Fig. 10. Probability of error detection for the Gaussian distributed host signal.

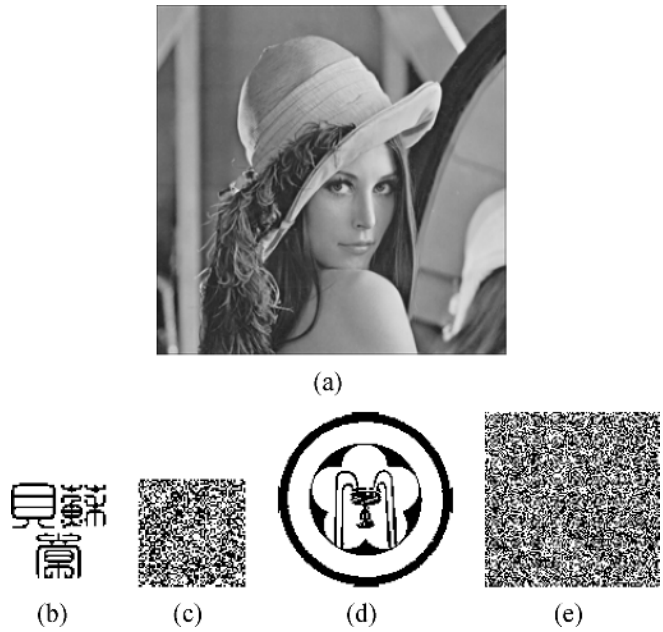

Fig. 11. Host and the watermark images: (a) Host image (256 × 256). (b) Watermark image (64 × 64). (c) Scrambled image of (b). (d) Watermark image (128 × 128). (e) Scrambled image of (d).

Fig. 12. Watermarked images for (a) $L = 16$ and (b) $L = 4$. ($\mathrm{PSNR} = 40$ dB).

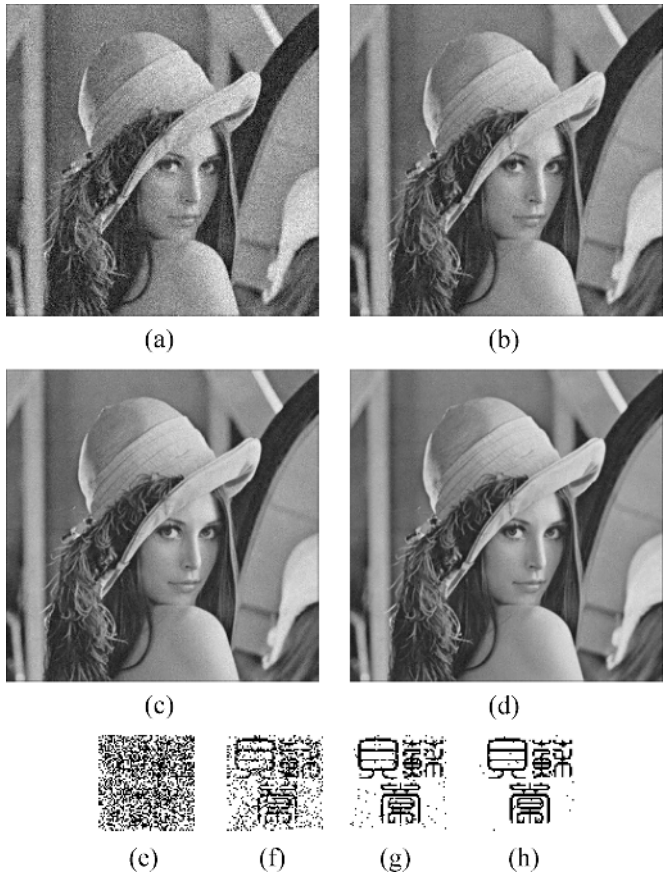

Fig. 13. Watermarked image ($L = 16$) is subject to the Gaussian noise attack such that the PSNR is (a) 25 dB, (b) 30 dB, (c) 35 dB, and (d) 38 dB, respectively. (e)–(h) Extracted watermark images of (a)–(d), respectively.

B. Embedding Bilevel Watermark Into Gray-Level Image

In this subsection, the host signal comes from a 256 × 256 gray-level image, shown in Fig. 11(a). As discussed in Section II-B, the capacity can be traded for reducing the embedding-induced distortion. Therefore, the host image is divided block-wise and each of the blocks forms a vector $\mathbf{x} \in \mathbb{R}^L$. Every vector conveys one bit of the watermark message. Fig. 11(b) and (d) shows the bilevel watermark images of 64 × 64 and 128 × 128 for the embedding capacities of 1/16 and 1/4 bits per sample, respectively. As discussed in Section I, both of the watermark images are scrambled to resist the cropping attack, and the $P_r\{m = 1\}$ and $P_r\{m = 0\}$ are equalized, as shown in Fig. 11(c) and (e). The embedding is conducted by quantizing the projection value of the host vector along a random direction. To let the embedding be unnoticed, the quantization step is adjusted to the extent that the PSNR of the watermarked image is 40 dB. Fig. 12(a) and (b) shows the watermarked images of the proposed scheme for $L = 16$ and 4, respectively.

The robustness test is first conducted by superimposing the Gaussian noise $n \sim \mathcal{N}(0, \sigma_n^2)$ on the watermarked image, where σ_n^2 is set to the extent that the PSNR of attacked watermarked image ranges from 25 to 38 dB. Figs. 13 and 14 show the attacked watermark and the extracted watermark images. Obviously, provided the noise does not deteriorate the watermarked image too much, the extracted watermark image is clearly discernible. The difficulty in recognizing the

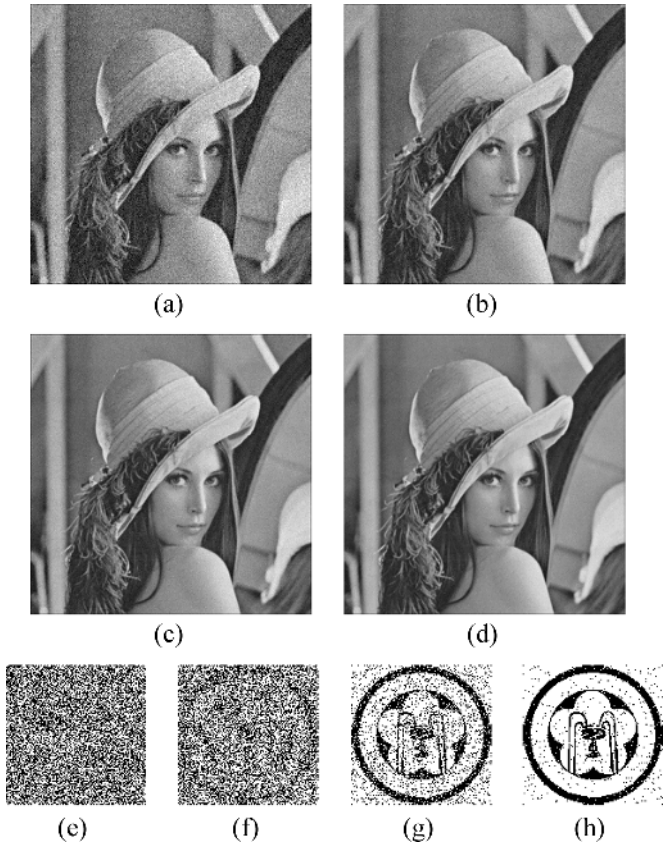

Fig. 14. Watermarked image ($L = 4$) is subject to the Gaussian noise attack such that the PSNR is (a) 25 dB, (b) 30 dB, (c) 35 dB, and (d) 38 dB, respectively. (e)–(h) Extracted watermark images of (a)–(d), respectively.

extracted watermark image reduces with decreasing distortion imposed on the watermarked image. Observing and comparing Figs. 13 and 14 can clarify the influence of L on robustness. The extracted watermark is still faintly visible when the watermarked image deteriorats to $\mathrm{PSNR} = 30$ dB in the case of $L = 16$. However, in the same scenario for $L = 4$, the extracted watermark image just appears to be a random pattern. Fig. 15(a) and (b) compares the robustness of the proposed scheme with that of other quantization-based watermarking schemes when the watermarked image is subjected to Gaussian noise attack. As in Section IV-A, the proposed scheme outperforms the other two quantization-based watermarking schemes.

The following compares the robustness of the proposed scheme and other quantization-based watermarking schemes under nonGaussian noise attack. Here, different levels of JPEG compression are applied to the watermarked image, because the JPEG standard is a common method of compressing the still images. Fig. 16 shows the error probabilities of the extracted watermark. The proposed scheme still outperforms the conventional quantization-based watermarking scheme when the watermarked image is subject to a high compression; and always outperforms the LUT approach.

In Figs. 15 and 16, the compared approaches are all blind watermarking methods except for the informed-SS of which performance, as aforementioned, was presented as the bound [9]. As regards the blind-SS, it inevitably suffers the interference from the host signal, which seriously degrades the robustness

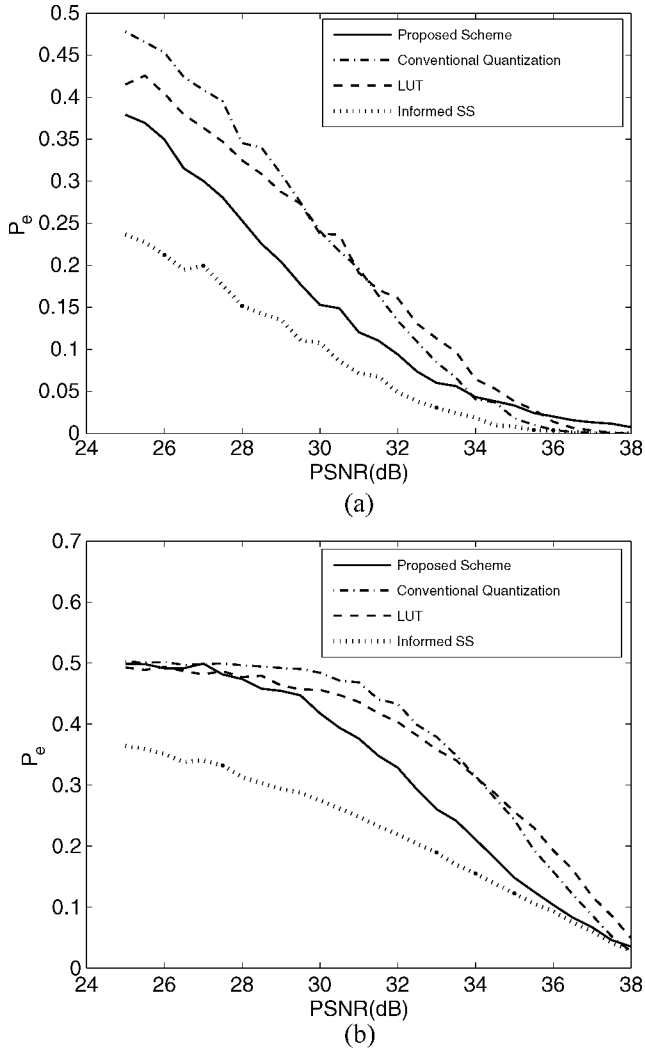

Fig. 15. Probability of error detection where the watermarked image is subjected to Gaussian noise attack. The PSNR of the watermarked image is 40 dB. (a) $L = 16$. (b) $L = 4$.

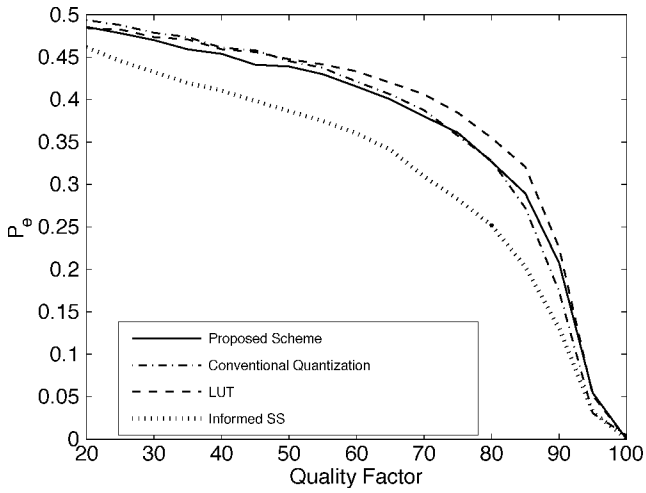

Fig. 16. Error probability of detection where the watermarked image is subjected to JPEG compression. The PSNR of the watermarked image is 40 dB.

performance as shown in Fig. 5, and thereby be excluded in those comparisons.

V. Conclusion

Based on the quantization approaches applied in compression or coding, the quantization-based watermarking schemes generally quantize the host value to the center of the quantization interval. However, this inheritance only considers the ability to resist attack, and hence the control of quantization error (the embedding-induced distortion in watermarking) resorts to adjusting the size of the quantization interval. This work presents that this is not necessarily true, yet the quantized value should be flexibly determined by the host value. Given tolerable embedding-induced distortion and the watermark message to be embedded, this work proposes a method of analytically determining the quantized value, and thus reaches a compromise between robustness requirements and the transparency of the watermarking scheme.

The simulation results demonstrate that the proposed scheme outperforms the other quantization-based watermarking schemes. In the scenario of inducing the same degree of distortion, the proposed scheme always has lower probability of error detection than the other schemes. Additionally, the proposed scheme has better nondisclosure of the quantization step owing to the difficulty of inferring the quantization step from the watermarked data.

References

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.

[2] C. T. Hsu and J. L. Wu, "Hidden digital watermarks in images," *IEEE Trans. Image Process.*, vol. 8, no. 1, pp. 58–68, Jan. 1999.

[3] S. C. Pei and J. H. Chen, "A robust blind image watermarking scheme based on vector quantization," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E87-A, no. 4, pp. 912–919, Apr. 2004.

[4] H. C. Huang, F. H. Wang, and J. S. Pan, "Efficient and robust watermarking algorithm with vector quantisation," *Electron. Lett.*, vol. 37, no. 13, pp. 826–828, Jun. 2001.

[5] A. Baştuğ and B. Sankur, "Improving the payload of watermarking channels via ldpc coding," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 90–92, Feb. 2004.

[6] B. MacQ, J. Dittmann, and E. J. Delp, "Benchmarking of image watermarking algorithms for digital rights management," *Proc. IEEE*, vol. 92, no. 6, pp. 971–984, Jun. 2004.

[7] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[8] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. San Mateo, CA: Morgan Kaufmann, 2002.

[9] H. S. Malvar and D. A. F. Florêncio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[10] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[11] M. Wu, "Joint security and robustness enhancement for quantization based data embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 831–841, Aug. 2003.

[12] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Process.*, vol. 66, no. 3, pp. 357–372, 1998.

[13] P. Bao and X. Ma, "Image adaptive watermarking using wavelet domain singular value decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 96–102, Jan. 2005.

[14] C.-H. Chang, Z. Ye, and M. Zhang, "Fuzzy-art based adaptive digital watermarking scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 65–81, Jan. 2005.

[15] R. M. Gray and J. T. G. Stockham, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1983.

[16] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, vol. 86, pp. 1064–1087, Jun. 1998.

Soo-Chang Pei (SM'89–F'00) was born in Soo-Auo, Taiwan, R.O.C., in 1949. He received B.S.E.E. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1970 and M.S.E.E. and Ph.D. degrees from the University of California, Santa Barbara, in 1972 and 1975, respectively.

From 1970 to 1971, he was an engineering officer in the Chinese Navy Shipyard. From 1971 to 1975, he was a Research Assistant at the University of California, Santa Barbara. He was the Professor and Chairman in the Electrical Engineering department of Tatung Institute of Technology and National Taiwan University from 1981 to 1983 and 1995 to 1998, respectively. Presently, he is the Dean of Electrical Engineering and Computer Science College and the Professor of Electrical Engineering department at National Taiwan University. His research interests include digital signal processing, image processing, optical information processing, and laser holography.

Dr. Pei received the National Sun Yet-Sen Academic Achievement Award in Engineering in 1984, the Distinguished Research Award from the National Science Council from 1990 to 1998, the Outstanding Electrical Engineering Professor Award from the Chinese Institute of Electrical Engineering in 1998, the Academic Achievement Award in Engineering from the Ministry of Education in 1998, the Pan Wen-Yuan Distinguished Research Award in 2002, and the National Chair Professor Award from Ministry of Education in 2002. He has been President of the Chinese Image Processing and Pattern Recognition Society in Taiwan from 1996 to 1998, and is a member of Eta Kappa Nu and the Optical Society of America (OSA). He became an IEEE Fellow in 2000 for contributions to the development of digital eigenfilter design, color image coding and signal compression, and to electrical engineering education in Taiwan.

Jun-Horng Chen (S'02–M'05) was born in Yu-Lin, Taiwan, R.O.C., in 1966. He received the B.S. degree in electronic engineering from National Taiwan University of Science and Technology, Taiwan, R.O.C., in 1991, and the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taiwai, R.O.C., in 1993 and 2005, respectively.

He is currently an Associate Professor and the Chairman of the Communication Engineering Department, Oriental Institute of Technology, Taipei, Taiwan, R.O.C. His research interests include digital image processing and wireless communication.